

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 17:08:19 ; Search time 20 Seconds  
(without alignments)  
1759.887 Million cell updates/sec

Title: US-10-060-848-3  
Perfect score: 1942  
Sequence: 1 MNYSKDAPEFVVSPKDAREF.....ALEELATKCDVQMFSYKRLK 366

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	1008	51.9	372	2	T20396	hypothetical prote
2	924	47.6	400	2	T18570	probable malate de
3	511	26.3	360	2	B71073	probable malate de
4	501	25.8	362	2	H75131	malate dehydrogena
5	465.5	24.0	339	2	S08981	malate dehydrogena
6	456.5	23.5	344	2	H64477	malate dehydrogena
7	407	21.0	341	2	H69027	malate dehydrogena
8	407	21.0	369	2	F96969	malate dehydrogena
9	390.5	20.1	337	2	E69852	malate dehydrogena
10	389.5	20.1	355	2	T20825	hypothetical prote
11	377.5	19.4	361	2	T46704	hypothetical prote
12	371	19.1	334	2	E95279	probable [imported
13	363.5	18.7	349	2	AB0568	ureidoglycolate de
14	360.5	18.6	349	2	C90701	ureidoglycolate de

*C. elegans*

{hypothetical prote  
probable malate de  
probable malate de  
malate dehydrogena  
malate dehydrogena  
malate dehydrogena  
malate dehydrogena  
malate dehydrogena  
malate dehydrogena  
hypothetical prote  
hypothetical prote  
probable [imported  
ureidoglycolate de  
ureidoglycolate de

*Super family  
malate dehydrogenase*

15	360.5	18.6	349	2	F85551	malate dehydrogena
16	356.5	18.4	349	2	D64783	malate dehydrogena
17	331.5	17.1	361	2	G90738	probable dehydroge
18	331.5	17.1	361	2	A85589	probable dehydroge
19	326.5	16.8	361	2	A64817	malate dehydrogena
20	269.5	13.9	332	2	A65157	hypothetical prote
21	268	13.8	349	2	I39535	probable L-lactate
22	266	13.7	334	2	D95361	probable malate de
23	261	13.4	345	2	AD3635	malate dehydrogena
24	257.5	13.3	332	2	C64165	hypothetical prote
25	255	13.1	346	2	AF2596	malate dehydrogena
26	255	13.1	346	2	H97378	probable l-malate
27	254	13.1	332	2	AD0979	probable carboxyli
28	241	12.4	345	2	AH3131	malate dehydrogena
29	241	12.4	346	2	C98156	probable L-malate
30	240.5	12.4	358	2	A98332	malate dehydrogena
31	239	12.3	345	2	C95873	probable malate de
32	227.5	11.7	309	2	AB2951	malate dehydrogena
33	227.5	11.7	334	2	B83488	probable L-malate
34	104.5	5.4	3300	2	D70575	probable PPE prote
35	100.5	5.2	370	2	A97235	RCC1 repeats prote
36	98.5	5.1	1145	2	G87284	hypothetical prote
37	98	5.0	829	2	E87305	TonB-dependent rec
38	95.5	4.9	382	2	T49762	hypothetical prote
39	95	4.9	543	2	S46098	probable acid-CoA
40	94.5	4.9	382	2	I39780	subtilisin (EC 3.4
41	94	4.8	512	2	A87270	hypothetical prote
42	94	4.8	2154	2	F83068	hypothetical prote
43	93.5	4.8	540	2	F96841	hypothetical prote
44	93	4.8	487	2	T49424	hypothetical prote
45	93	4.8	1101	2	T21062	hypothetical prote

# ALIGNMENTS

## RESULT 1

T20396

hypothetical protein F36A2.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C;Accession: T20396; T21836

R;Lennard, N.

submitted to the EMBL Data Library, June 1997

A;Reference number: Z19266

A;Accession: T20396

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-372 <WIL>

A;Cross-references: EMBL:Z96047; PIDN:CAB09417.1; GSPDB:GN00019; CESP:F36A2.3

A;Experimental source: clone DY3

R;Lennard, N.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19476

A;Accession: T21836

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-372 <WI2>  
A;Cross-references: EMBL:Z81077; PIDN:CAB03073.1; GSPDB:GN00019; CESP:F36A2.3  
A;Experimental source: clone F36A2  
C;Genetics:  
A;Gene: CESP:F36A2.3  
A;Map position: 1  
A;Introns: 21/3; 53/2; 79/1; 114/3; 147/3; 198/3; 249/1; 284/3; 353/3  
C;Superfamily: malate dehydrogenase ylbC

Query Match 51.9%; Score 1008; DB 2; Length 372;  
Best Local Similarity 54.6%; Pred. No. 1.2e-72;  
Matches 194; Conservative 48; Mismatches 111; Indels 2; Gaps 2;

```

Qy      6 DAPEFVVSPKDAREFVVKCMQTVGTSPDHAGQLADLLLDADLVGHYSHGLNRLHIYVDD- 64
      : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     11 ETDEIVISKEKLDSEFVLECLAKAGCTGDHAQQLAETLLCSDYRGHYSHGINRLHIYVHDL 70

Qy     65 VKNGVKGNGVPKVLKQKGGTAWVDGENLLGAVVGNFCTDLAIKLAKEFGVAWVVTKNSNH 124
      : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     71 MMKSTAVTGTTPQVLKSKGSTAWVDGNLLGPPVGNFCMQLAVEKAKEFGIGWVVCNRNSNH 130

Qy    125 YGACQHYTKKIANAGMVGMSFTNTSPLMFPCRSEIGLGTNPLSCCVNSEKTGDSFLLDM 184
      : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    131 FGIAGWYADFACRNLVGMFAFTNTSPCVFPTGSREKSLGSNPI-CMAAPGMEGDSFFLDM 189

Qy    185 ATTTVALGKVELADCRGKTQIPSTWGADSKGNPSTDTQVVLHGGGLLPLGGIEETGSYKG 244
      | : | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db    190 ASTTVAYGKIEVVDRKGETYIPGSWGADKNGDETHNPKEVLDGGGLQPLGGSEITGGYKG 249

Qy    245 TGLSMMGELFCGILAGSSFGKNVRLWGQSHKAADNGQCFVAIDQECFAPGFAPRLQQFLD 304
      | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db    250 TGLCMMVEVLCGIMGGSAFGKNIRQWQTTSKTADLGQCFVAIDPECFAPGFNSRLQEFCD 309

Qy    305 ETRNLKPISEEEKPVLVPGDPERMNTTEYSQKAGGLVYQEGQIKALEELATKCDVQM 359
      | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db    310 ETRNLNPINPSRPPQVPGDPERAHMNMCDLGGIVYKKKQLDHLKKNLADRLGVIM 364

```

## RESULT 2

T18570

probable malate dehydrogenase (EC 1.1.1.37) - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C;Accession: T18570; T20833

R;Ainscough, R.

submitted to the EMBL Data Library, December 1998

A;Reference number: Z18979

A;Accession: T18570

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-400 <WIL>

A;Cross-references: EMBL:AL033535; PIDN:CAA22133.1; CESP:VF13D12L.3

A;Experimental source: clone VF13D12L

R;Coles, L.

submitted to the EMBL Data Library, April 1995

A;Reference number: Z19330

A;Accession: T20833

A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-400 <WI2>  
A;Cross-references: EMBL:Z49127; PIDN:CAA88951.1; GSPDB:GN00020; CESP:VF13D12L.3  
A;Experimental source: clone F13D12  
C;Genetics:  
A;Gene: CESP:VF13D12L.3  
A;Map position: 2  
A;Introns: 53/3; 96/1; 179/3; 230/3; 276/1; 315/3; 347/3; 384/3  
C;Superfamily: malate dehydrogenase ylbC  
C;Keywords: oxidoreductase

Query Match 47.6%; Score 924; DB 2; Length 400;  
Best Local Similarity 53.8%; Pred. No. 6.5e-66;  
Matches 191; Conservative 48; Mismatches 110; Indels 6; Gaps 5;

```

Qy      7 AP-EFVVSPKDAREFVVVKCMQTVGTSPDHAGQLADLLLDADLVGHYSHGLNRLHIYVDDV 65
      || | ||: : : |:|:| | | : | | || :|:| : | ||||| | | :| | :
Db     43 APEESVVAKDEMKRFMVECMTKVGATESHATQLALVLLEGDIRGHYSHGLNRLDMYVRDI 102

Qy     66 KNGV-KGNGVPKVLKQKGGTAWVDGENLLGAVVGNFCTDLAIKLAKEFGVAWVVTKNSNH 124
      : | ||:| | :|:| | ||||| | || | ||||| | ||:| | : | || | |||
Db    103 EQNVCKGDGEPIILKEKAGTAWVDGNNLLGPVVGNFCDLAIEKAKNAGIGWVVAKGSNH 162

Qy    125 YGACQHYTKKIANAGMVGMSFTNTSPLMFPCRSEIGLGTNPLSCCVNSEKTG-DSFLLD 183
      || | | : | |:| | ||||:| | | | : | ||||:| : : | | |||:| |
Db    163 YGIAGWYALRAMKKGMLGMSMTNTSPISFPTRSAVPALGTNPIS--LAAPGTGDDSFVLD 220

Qy    184 MATTTVALGKVELADCRGKTQIPSTWGADSKGNPSTDTQVVLHGGGLLPLGGIEETGSYK 243
      ||:| |||:| |||| | : :| :| | :| | ||:| ||||| |||:| :| ||
Db    221 MASTTVAIGKVELA-ARKENPVPLSWGVEGGKETTDPTKVLYGGGLLPLGGVEVSGGYK 279

Qy    244 GTGLSMMGELFCGILAGSSFGKNVRLWGQSHKAADNGQCFVAIDQECFAPGFAPRLQQFL 303
      | ||| | |:| ||||| : :| :| | : | | ||||| | | ||||| ||| :
Db    280 GYGLSSMIEIFCGILAGAHWGPHVRKWMSTKSEADLGQCFVAIDPEAFAPGFADRLQDFM 339

Qy    304 DETRNLKPISEKPVLPVPGDPERMNTSEYSQKAGGLVYQEGQIKALEELATKCDVQ 358
      | | | | | | | | : : : | : | : | : | : | | |
Db    340 QTMRALPTSSPSFKVEVAGDMERRHEALVEQLGGIPYHKNQITFVNDLAAKLGVK 394

```

# RESULT 3

B71073

probable malate dehydrogenase - *Pyrococcus horikoshii*

C;Species: *Pyrococcus horikoshii*

C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000

C;Accession: B71073

R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H. DNA Res. 5, 55-76, 1998

A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3.

A;Reference number: A71000; MUID:98344137; PMID:9679194

A;Accession: B71073

A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-360 <KAW>  
A;Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30380.1; PID:g3257697  
A;Experimental source: strain OT3  
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C;Genetics:  
A;Gene: PH1277  
C;Superfamily: malate dehydrogenase ylbC

Query Match 26.3%; Score 511; DB 2; Length 360;  
Best Local Similarity 36.9%; Pred. No. 4.6e-33;  
Matches 128; Conservative 66; Mismatches 139; Indels 14; Gaps 8;

Qy	10	FVVSPKDAR-EFVVVKCMQTVGTSPDHAGQLADLLLDADLVGHYSHGLNRLHIYVDD-VKN	67
		::      ::: :   :   :   ::      .    :        :	
Db	11	YIRVPKDRLFSFIVRVLTCLKGVPEEDAKIVADNLVMADLRGVESHGVQRLKRYVDGIISG	70
Qy	68	GVKNGVGPVKVLKQKGGTAWVDGENLLGAVVGNFCTDLAIKLAKEFGVAWVVTKNSNHYGA	127
		: : :::   :  :              ::  :  : :	
Db	71	GVNLHPKIRVIREGPSYALIDGDEGLGQVVGYSMKLAIKKAKDTGIGIVIARNSNHYGI	130
Qy	128	CQHYTKKIANAGMVGMSFTNTSPLMFPCRSEIGLGTNPLSCCVNSEKTGDSFLLDMATT	187
		:      : :    :   :         :: : :      :	
Db	131	AGYYALMAAEEGMIGISMTNSRPLVAPTGGIERILGTNPIALAAPT--DKPFLDMATS	188
Qy	188	TVALGKVELADCRGKTQIPSTWGADSKGNPSTDTQVVLHGGGLLPLGGIEE-TGSYKGTG	246
		:  : : :        : :   :   :             :	
Db	189	VVPIGKLEVYRRKGK-DIPEGWAINREGNITTKVEEVFNKGALLPLGGFGELLGGHKGYG	247
Qy	247	LSMMGELFCGILAGSSFGKNVRLWGQSHKAADNGQCFVAIDQECFAP--GFAPRLQQFLD	304
		:  ::    :  ::    :     ::  :            : :   ::	
Db	248	LSLMVDILSGILSGGTWSKYVK--NTSEKGSNVCHFFMVIDIEHFIPLEEFKEKISQMIE	305
Qy	305	ETRNLKPISEKPVLPVPGDPERMNTSEYSQKAGGLVYQEGQIKALEEL	351
		:: :   : : :  : :       : :	
Db	306	EIKSSRKHPEFERIWIHGEKGFLTMETRLKLGIPYR---KVLEEL	348

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 17:06:33 ; Search time 18 Seconds  
(without alignments)  
956.210 Million cell updates/sec

Title: US-10-060-848-3  
Perfect score: 1942  
Sequence: 1 MNYSKDAPEFVVSPKDAREF.....ALEELATKCDVQMFSYKRLK 366

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query		Length	DB	ID	Description
	Score	Match				
1	511	26.3	360	1	MDH_PYRHO	O59028 pyrococcus
2	501	25.8	362	1	MDH_PYRAB	Q9v0d5 pyrococcus
3	465.5	24.0	339	1	MDH_METFE	P16142 methanother
4	456.5	23.5	344	1	COMC_METJA	Q58820 methanococc
5	407	21.0	341	1	COMC_METTH	O27273 methanobact
6	390.5	20.1	337	1	YJMC_BACSU	O34736 bacillus su
7	360.5	18.6	349	1	ALLD_ECO57	P58408 escherichia
8	356.5	18.4	349	1	ALLD_ECOLI	P77555 escherichia
9	331.5	17.1	361	1	YBIC_ECO57	P58409 escherichia
10	326.5	16.8	361	1	YBIC_ECOLI	P30178 escherichia
11	269.5	13.9	332	1	YIAK_ECOLI	P37672 escherichia
12	268	13.8	349	1	LDH_ALCEU	Q07251 alcaligenes
13	258	13.3	332	1	YC56_PASMU	Q9clh5 pasteurella
14	257.5	13.3	332	1	YIAK_HAEIN	P44995 haemophilus
15	227.5	11.7	334	1	YC52_PSEAE	Q9i492 pseudomonas
16	99	5.1	940	1	GBR2_RAT	O88871 rattus norv
17	97	5.0	379	1	RECA_STRTR	Q9evv7 streptococc
18	97	5.0	655	1	ACDV_BOVIN	P48818 bos taurus

19	96.5	5.0	383	1	RECA_STRMU	P27624	streptococc
20	96.5	5.0	860	1	VG12_BPB03	Q37893	bacterioph
21	96	4.9	378	1	RECA_STRP3	Q8k5k0	streptococc
22	95	4.9	543	1	FAT2_YEAST	P38137	saccharomyc
23	92.5	4.8	363	1	EFTU_GYMST	O63930	gymnochlora
24	92.5	4.8	1017	1	KCH4_HUMAN	Q9uq05	homo sapien
25	92	4.7	941	1	GBR2_HUMAN	O75899	homo sapien
26	92	4.7	1156	1	PHYB_SOYBN	P42499	glycine max
27	91.5	4.7	916	1	GYRA_NEIGO	P48371	neisseria g
28	91	4.7	378	1	RECA_STRP8	Q8nz30	streptococc
29	91	4.7	378	1	RECA_STRPY	Q59942	streptococc
30	90	4.6	494	1	AMYR_DROPS	O18552	drosophila
31	90	4.6	900	1	IF2_MYCTU	P71613	mycobacteri
32	89.5	4.6	739	1	PURL_LACLC	Q9zb06	lactococcus
33	88.5	4.6	424	1	GATD_METAC	Q8tr66	methanosarc
34	87.5	4.5	333	1	YRP5_IRV6	P18309	chilo iride
35	87.5	4.5	575	1	FLA2_CAMJE	P22251	campylobact
36	87.5	4.5	728	1	SPE1_PEA	Q43075	pisum sativ
37	87.5	4.5	997	1	YPX2_CAEEL	Q20256	caenorhabdi
38	87.5	4.5	1023	1	RT11_ACTPL	P55128	actinobacil
39	87	4.5	534	1	PROA_VIBAL	P16588	vibrio algi
40	87	4.5	536	1	SPM1_MAGGR	P58371	magnaporthe
41	86.5	4.5	381	1	RECA_STRPA	O85502	streptococc
42	86.5	4.5	387	1	YF20_METJA	Q58915	methanococc
43	86.5	4.5	394	1	EFTU_BORBU	P50062	borrelia bu
44	86.5	4.5	575	1	FLB2_CAMJE	P22252	campylobact
45	86.5	4.5	978	1	PM14_CHLPN	Q9z895	chlamydia p

#### ALIGNMENTS

##### RESULT 1

##### MDH\_PYRHO

ID MDH\_PYRHO STANDARD; PRT; 360 AA.  
AC O59028;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Malate dehydrogenase (EC 1.1.1.37).  
GN MDH OR PH1277.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
RT thermophilic archaebacterium, Pyrococcus horikoshii OT3.";

